



**University of  
Zurich<sup>UZH</sup>**

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2018

---

## **Quantification of the Voicescape: A Person-centric Approach to Describing Real-life Behaviour Patterns - A Case Study Comparing Two Age Groups**

Londral, Ana ; Demiray, Burcu ; Cheetham, Marcus

**Abstract:** The human voice is a fundamental part of the everyday auditory environment. A measure of all voice activity that a person produces or perceives in the environment, i.e., the person's voicescape, might provide an informative, low cost, ecologically valid, and person-centric approach to characterizing patterns of socially-relevant behaviour in real life. In this paper, we use the measure ratio of voice activity (rva) and present results of data acquired from N=20 subjects of 2 different age groups as they engaged in their usual daily life activities over 4 consecutive days. The data show no differences in total voice activity but significant between-group differences in its daily distribution. We propose that measurement of the voicescape can, even without knowledge of specific voice sources, serve as a useful indicator of person- or group specific activity patterns for purposes of describing significant aspects of variation and within- and between-group differences in patterns of everyday behaviour and, potentially, for identifying change in patterns that have health-related implications. Future work will target automatic detection and identification of voice sources and the use of privacy-preserving processing methods.

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-152481>

Conference or Workshop Item

Published Version

Originally published at:

Londral, Ana; Demiray, Burcu; Cheetham, Marcus (2018). Quantification of the Voicescape: A Person-centric Approach to Describing Real-life Behaviour Patterns - A Case Study Comparing Two Age Groups. In: 11th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2018), Funchal, Madeira, Portugal, 19 January 2018 - 21 January 2018. s.n., 311-314.

# Quantification of the voicescape: A person-centric approach to describing real-life behaviour patterns.

## *A case study comparing two age groups*

Ana Londral,<sup>1,2</sup> Burcu Demiray,<sup>3,4</sup> Marcus Cheetham<sup>1,3</sup>

<sup>1</sup>. Department of Internal Medicine, University Hospital Zurich, Zurich, Switzerland

<sup>2</sup>Institute of Molecular Medicine, University of Lisbon, Portugal

<sup>3</sup>. University Research Priority Program “Dynamics of Healthy Aging”, University of Zurich, Zurich, Switzerland.

<sup>4</sup>. Division of Gerontopsychology and Gerontology, Department of Psychology, University of Zurich, Zurich, Switzerland

**Keywords:** Speech Signal Processing, Voicescape, Person-centric, Behaviour analysis, Ageing

**Abstract:** The human voice is a fundamental part of the everyday auditory environment. A measure of all voice activity that a person produces or perceives in the environment, i.e., the person’s *voicescape*, might provide an informative, low cost, ecologically valid, and person-centric approach to characterizing patterns of socially-relevant behaviour in real life. In this paper, we use the measure *ratio of voice activity (rva)* and present results of data acquired from N=20 subjects of 2 different age groups as they engaged in their usual daily life activities over 4 consecutive days. The data show no differences in total voice activity but significant between-group differences in its daily distribution. We propose that measurement of the voicescape can, even without knowledge of specific voice sources, serve as a useful indicator of person- or group specific activity patterns for purposes of describing significant aspects of variation and within- and between-group differences in patterns of everyday behaviour and, potentially, for identifying change in patterns that have health-related implications. Future work will target automatic detection and identification of voice sources and the use of privacy-preserving processing methods.

## 1 INTRODUCTION

The human voice is a fundamental part of a person’s everyday auditory environment. The combination of all voice activities that a person produces or perceives within the auditory environment may be referred to as the person’s *voice soundscape* (cf. Schafer 1994[1977]) or simply *voicescape*. Even at a low-level of granularity, the voicescape provides a potentially informative means to exploring and characterizing patterns of socially-relevant behaviour from a person-centric perspective in the natural setting.

Real-life patterns of behavioural activity are of longstanding interest (Fahrenberg et al, 2007), more recently in the field of aging research. Age-related research of everyday behaviours has focussed on physical activity, with growing interest in mobility, social context, and time-location patterns (e.g., Khusainov et al, 2013; Berke et al, 2011; Yang & Hsu 2010). But as the physical and social environment of the speaker changes due to the impact of aging on normal functioning in daily life (Wahl & Lang, 2004), quantitative and qualitative aspects of the voicescape may be expected to change, too.

In the present paper, we analysed all voice activity of 2 groups of 20 younger and older healthy subjects in sound samples acquired unobtrusively using a wearable device while subjects engaged in their daily activities. For analysis, we used the feature ratio of voice-to-nonvoice activity (*rva*) in each sound sample as an indicator of activity in the voicescape.

This paper describes the tool that was developed to extract and quantify voice activity in the samples, presents the results based on *rva*, and demonstrates the visualization tool developed to present the results. The results show different patterns of activity in the voicescape of the two age groups.

We suggest that low-level measures of the voicescape (e.g. *rva*, sound versus silence, noise exposure) can play a useful role in describing and bringing attention to significant aspects of variation and within- and between-group differences in cross-sectional or longitudinal patterns of everyday behaviour. We discuss the potential application of a person's voicescape in studies of health aging and e.g. for alerting health professionals to unexpected health-relevant changes in habitual patterns of behavioural activity that may escape self-report.

## 2 METHODOLOGY

### 2.1 Dataset

For this study, we used a data set that was collected over 4 consecutive days, using the *Electronically Activated Recorder (EAR)* (Mehl et al, 2001) as a method for naturalistic observation in daily life. Each participant used a mobile phone and app to randomly activate the microphone to record 30 seconds of sound (on average 4

samples/hour). Each 30s sample was stored as a wav. file. The recordings were active from 6:00 to 00:00. The protocol is otherwise as described e.g. in Mehl & Pennemaker (2003).

We analysed 20 participants from this dataset: 10 in the age group of the young (Y) and 10 in the age group of the elderly (E).

### 2.2 Extraction of voice activity

Praat was used for voice activity detection and segmentation. This tool is an opensource software for speech signal processing (Boersma and Weenink 2007).

From the process of segmentation, the total amount of voice per each sample was calculated as the sum of the duration of all the voiced segments. According to equation (Equation 1), the *ratio of voice activity (rva)* is defined for each sample as the percentage of voice activity in the total 30-sec sample.

$$rva = \frac{\sum \Delta t(vseg_i)}{30} \quad (\text{Equation 1})$$

### 2.3 Grouping and Resampling

For each subject, we re-sampled the data to 2h periods, and automatically calculated the average values of *rva* of all the samples recorded in each of the following periods: {6:00-8:00; 8:00-10:00; 10:00-12:00; 12:00-14:00; 14:00-16:00; 16:00-18:00; 18:00-20:00; 20:00-22:00; 22:00-24:00}.

## 3 RESULTS

A total of 55.7 hours of sound were recorded and 23.07 hours of voice activity detected in the voicescape of the 20 subjects (described in Table 1), The *rva* per sample is represented in Figure 1,

indicating that most samples have either very low voice activity ( $rva < 2\%$ ) or very high voice activity ( $rva > 98\%$ ), independently of the age group.

Table 1: Description of the dataset.

<b>Total of persons</b>	<b>20</b>
Age groups	Old /Young
Average amount of sound recorded per subject (hours)	$3.1 \pm 0.9$
Average voice detected per subject (min)	$76.9 \pm 23.2$
Average voiced detected per age group	$74.5 \pm 27.1$ (O) $79.4 \pm 18.1$ (Y)

A visual representation of the  $rva$  of a subject is presented in this paper as a tool for visual behaviour analysis. As depicted in Figure 2, the representation of voice activity for two subjects allows identification of differences in the voice environment that may be related to age characteristics. From Figure 2, it is clear that the subject in the group Y has more voice activity (a pick of  $rva$  in the period 20:00-22:00), while the subject in group O has clearly less voice activity in the late evening period. Another interesting aspect to note from these graphics is that the  $rva$  for the subject in group O is higher by the end of the morning period and by beginning of evening period. This fact is probably related to lunch and dinner time, when this subject may have the most social moments.

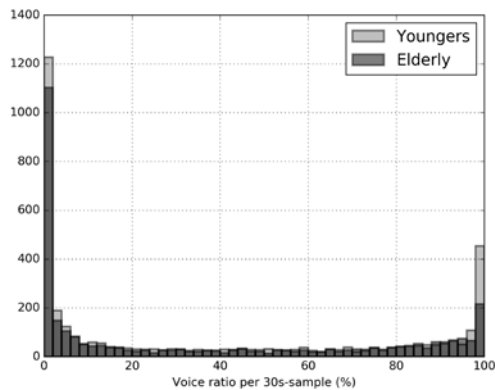


Figure 1: Histogram containing the  $rva$  per sample.

Statistical analysis of  $rva$  and different periods of the day indicate significant differences ( $p < 0.001$ ,  $\eta^2 = 0.02$ ) between age groups and time periods. Visual inspection of Figure 4 indicates a time period in which there are strong between-group differences in voice activity at 14:00-16:00 p.m. At this time, voice activity of subjects in the older group decreases drastically. On the other hand,  $rva$  at late evening is considerably higher for the younger group.

Interestingly, when considering the total  $rva$  independently of time of day, the voice activity is similar in both groups, as depicted in the histogram of Figure 1. While this suggests the same overall amount of voice activity for the subjects independently of the age group, the sources of voice activity may differ between groups.

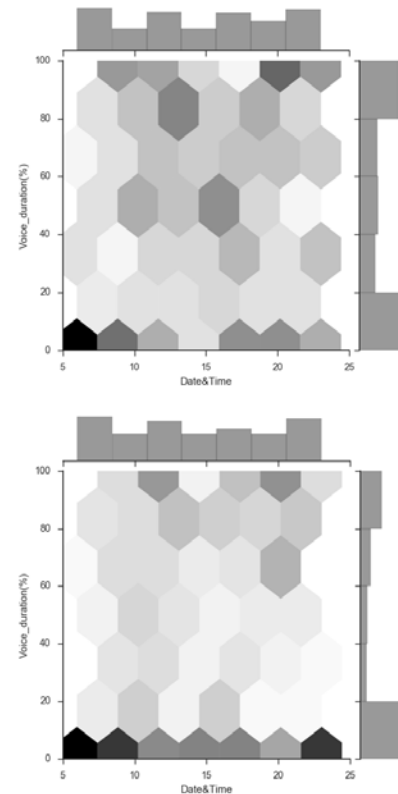


Figure 2: Graphical representation of voice activity of two subjects: Upper panel shows a subject from the younger group Y and the lower panel from the older group O.

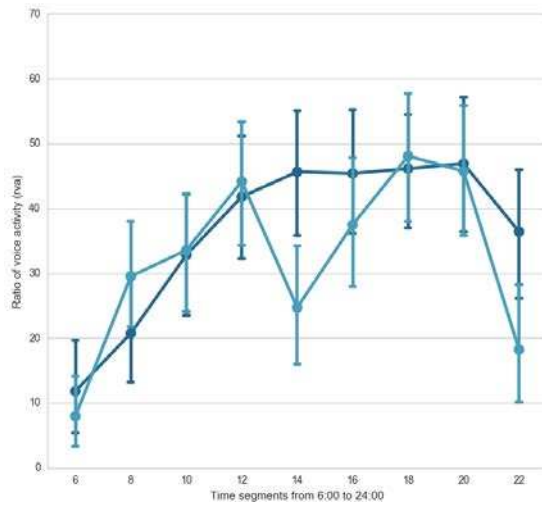


Figure 3: Mean *rva* in each time segment, separated by age groups. Vertical bars represent the confidence interval (95%).

Automatic classification of voice sources was not applied in this study. But manual auditory inspection was performed in random samples containing high *rva* (>90%). This showed that voice activity in older persons is mostly related to the TV, whereas in younger subjects voiced backgrounds (e.g. classroom or restaurant) dominate the voicescape.

### 3 CONCLUSIONS

In this paper, we propose the use of voice activity in the voicescape as a potentially informative, low cost, ecologically valid, person-centric approach to characterizing patterns of socially-relevant behaviour. In contrast to most studies in speech processing, where voice is recorded in controlled environments, this study used sounds recorded in the self-selected natural setting of the subjects.

The ratio of voice activity was analysed as the percentage of voiced segments in 30s-samples over 4 consecutive days to map daily patterns of voice activity. This

paper presents a case study with two age groups. We observed no overall difference in *rva* of younger and older adults in the natural setting and a common pattern of either extremely low or high *rva* in the samples, such that voice activity was either very low ( $rva < 2\%$ ) or very high ( $rva > 90\%$ ). Across the day, however, there were significant differences in voice activity in two specific time periods. Voice activity was lower in the older group over the midday period and late evening.

We conclude that voice activity present in a person's soundscape can, even without knowledge of specific voice sources, serve as an indicator of person- or group specific behavioural patterns for purposes of exploring significant areas of further research interest. This approach might be used to examine associations between health-related factors and patterns of habitual social-behavioural activity and to indicate deviations in habitual patterns of behaviour that may escape self-report but have health-related implications. Future work will target automatic detection and identification of voice sources. With a focus on the voicescape and voice sources, this work can be conducted using privacy-preserving processing methods (e.g., Glackin et al., 2017).

### ACKNOWLEDGEMENTS

This work was funded by the Department of Internal Medicine, University Hospital Zurich, Zurich, Switzerland and the University Research Priority Program "Dynamics of Healthy Aging", University of Zurich.

## REFERENCES

- Berke, E. M., Choudhury, T., Ali, S., & Rabbi, M. (2011). Objective Measurement of Sociability and Activity: Mobile Sensing in the Community. *Annals of Family Medicine*, 9(4), 344–350.
- Blake, E. C., & Cross, I. (2014). The Acoustic and Auditory Contexts of Human Behavior. *Current Anthropology*, 56 (1), 81-103. The University of Chicago Press Wenner-Gren Foundation for Anthropological Research
- Boersma, P., & D. Weenink. (2007). 'Praat: Doing Phonetics by Computer (Version 4.5.) [Computer Program]'. Retrieved from [Http://www.praat.org/](http://www.praat.org/) 5 (9/10): 341–45.
- Davies, William J. 2010. *The Acoustic Environment*. Oxford University Press.
- Fahrenberg, Jochen, Michael Myrtek, Kurt Pawlik, Meinrad Perrez., 2007. Ambulatory Assessment - Monitoring Behavior in Daily Life Settings: A Behavioral-Scientific Challenge for Psychology. *European Journal of Psychological Assessment* 23 (4): 206–13.
- Glackin, C., Chollet, G., Dugan, N., & Rajaraja, M. (2017). Privacy preserving encrypted phonetic search of speech data," 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, pp. 6414-6418.
- Ludlow, Christy L., 2015. Central Nervous System Control of Voice and Swallowing. *Journal of Clinical Neurophysiology: Official Publication of the American Electroencephalographic Society* 32 (4). NIH Public Access: 294–303.
- Khusainov, R., Azzi, D., Achumba, I. E., & Bersch, S. D. (2013). Real-Time Human Ambulation, Activity, and Physiological Monitoring: Taxonomy of Issues, Techniques, Applications, Challenges and Limitations. *Sensors (Basel, Switzerland)*, 13(10), 12852–12902.
- Mehl, M.R., Pennebaker, J.W., Crow, D.M. et al. (2001). *Behavior Research Methods, Instruments, & Computers*, 33; 517.
- Mehl MR, Pennebaker JW (2003) The sounds of social life: A psychometric analysis of students' daily social environments and natural conversations. *J Pers Soc Psychol.*, 84(4):857-70.
- Muaremi, A., Bert Arnrich, and Gerhard Tröster. 2013. 'Towards Measuring Stress with Smartphones and Wearable Devices During Workday and Sleep'. *BioNanoScience* 3 (2): 172–83.
- Narayanan, S., Panayiotis G., 2013. 'Behavioral Signal Processing: Deriving Human Behavioral Informatics from Speech and Language'. *Proceedings of the IEEE* 101 (5): 1203–33.
- Schafer, R. Murray. 1994 (1977). *The Soundscape: Our Sonic Environment and the Tuning of the World*. Rochester,. Vermont: Destiny, 1994 (1977). pp. 293.
- Wahl, H.-W. & Lang, F. (2004). Aging in context across the adult life course: Integrating physical and social environmental research perspectives. In H.-W. Wahl, R. Scheidt & P. Windley (eds.), *Aging in Context: Socio-Physical Environments (Annual Review of Gerontology and Geriatrics 23)* (pp. 1–34). New York: Springer
- Yang, C.-C., & Hsu, Y.-L. (2010). A Review of Accelerometry-Based Wearable Motion Detectors for Physical Activity Monitoring. *Sensors (Basel, Switzerland)*, 10(8), 7772–7788.